BladeRunner from Penguin Computing — A Turnkey Open Source Linux Clustering Solution

Analyst: David Reine

Management Summary

With the advent, and acceptance, of *Linux* as the open source operating system, and the increased availability of Linux applications for all size operations, **more companies than ever before are interested in Linux clusters as a consolidation tool and as a vehicle to install low-cost, scale-out solutions to provide greater computational power for High Performance Computing (HPC) problems within traditional technical and data mining environments**. Clusters have been used before in both open systems and proprietary environments to combine and unify the processing capability of independent servers via software or networking, such as *IBM's HA/CMP*, *Microsoft's Cluster Server (MSCS)* or *HP's Alpha* clusters. Now open source clustered systems are available to the open source community through the Beowulf Project, which started in 1994.

Beowulf Clusters are scalable performance clusters based on commodity hardware with the Linux operating system running as the underlying open source software infrastructure. The data center staff can improve performance proportionally by simply adding low-cost servers to the cluster. These commodity systems can be any of a number of standalone compute nodes using standard technology, such as SCSI, Ethernet, and IDE, with as few as two networked servers each running Linux and sharing a file system, or as complex as 1024 nodes over a high-speed, low-latency network. The original goal to prove that applications written for high-end cluster environments could run on low-cost clusters was quickly accomplished. It was replaced by a new goal: **to make a scalable cluster easy to install and manage for a non-technical staff**. That banner was picked up and carried forward by Scyld Software, founded in 1998 as the original pioneer in Linux clustering with the clear understanding that compute intensive applications require scalable yet robust computing environments. While Linux clusters provide compelling value for high performance computing (HPC), deploying and managing these systems can be very complex for system administrators. *Scyld Beowulf*, the company's robust and scalable software platform, was designed specifically to make Linux clusters easy!

Now there is an integrated solution to make that task even easier. Penguin Computing, founded in 1998 and dedicated to the Linux Operating System (a big surprise for a company named Penguin) has introduced a turnkey cluster solution based upon their *BladeRunner Cluster-in-a-box* platform and the Scyld Beowulf clustering software to provide a scalable platform for both consolidation and commercial computing environments. In order to find out how Penguin can help to reduce the total cost of ownership (TCO) for your computer operation, please read on.

The Linux Clustering Environment

There have always been ample clustering solutions for the larger enterprise with the resources, both money and people, to implement the complex solution available from vendors of proprietary clusters. There are, however, a great many companies that have been classified as small and medium businesses, or SMBs, who do not have the resources needed to make it happen. This category however spans too broad a range of businesses to be able to accurately position any given system solution. Companies in the SMB category could range anywhere from 10 employees to over a 1000. The Clipper Group has chosen to redefine the nomenclature for this category[1]. Instead of SMB, we address the Small-Scale (SSO), Mid-Scale (MSO), and Large-Scale (LSO) Operations, with the MSO represented by companies with more than a handful of servers, but less than many dozens, more than a few dozen employees, but not more than several hundred. **The MSO and the LSO looking to reduce their total cost of ownership for IT resources are the ideal candidates for Linux clustering**.

Initially, Linux was restricted in its target audience to universities and enterprises with high-performance applications. There it was nurtured by a devout following of dedicated administrators who followed in the footsteps of UNIX gurus from past decades. It was not broadly accepted within the Fortune 500 community as a platform for mission-critical applications. No one was willing to use Linux as the operating system of choice for a "bet-your-business" solution. Times have changed! **Today, Linux may not be the operating system of choice in every data center, but you will find Linux in every data center, even if it is in a development and test environment.**

Linux servers provide every CIO with the key factors he is looking for in upgrading an old application or introducing a new

one: a competitively priced, standards based, off-the-shelf platform with a growing following within the ISV community. ISVs value portability between servers because it reduces development costs and increases profits. UNIX was introduced to the enterprise community in the '80s with that same claim. Unfortunately, the data center did not acquire that flexibility between IBM's *AIX*, HP's *HP-UX*, Sun's *Solaris*, Digital's *Alpha*, and a myriad of other pretenders who fell by the wayside. They ended up installing a variety of proprietary UNIX solutions, each requiring a development port and the maintenance of multiple iterations of source code.

What UNIX did provide, and what the enterprise of the 20th century demanded, was a complete, robust, and scalable environment that had high availability (HA) and proven reliability, an environment that companies offering Linux solutions today must also address. However, the Linux platform today must also address the variables that UNIX did not. It must be easy to install, easy to use, easy to maintain, and most of all, easy to expand. It needs to be low-cost and it has to be fully supported by a staff of, if not expert, at least competent, engineers and administrators.

Linux has traveled a long distance in a relatively short period. No longer is it restricted to high-performance, technical, or academic arenas. Today, Linux can be found on PCs and mainframes. It is available on scale-up SMP servers, such as the HP Superdome, IBM's p595, Sun's E25K, and even on IBM's zSeries mainframe. It can be found on scale-out systems, clusters, from these same vendors, and more. More importantly, Linux has moved beyond academia. It can be found today growing in HPC environments, the Fortune 500 space where management, provisioning, and monitoring are valued highly, and in architectures that require the reliability only available in high-availability computing clusters. All indications are that the HPC and HA segments of the Linux community are growing the fastest, with expectations that they will represent almost half of all Linux servers by the end of 2006.

---

[1] See *The Clipper Group Captain's Log* dated December 14, 2004, entitled *Why "SMB" is a Meaningless Acronym - Trying to Define the "Middle"* and available at http://www.clipper.com/publications/TCG2004096.pdf.

Linux clusters are now available from a wide-range of smaller companies, as well as the mainstream trio. Several vendors are dedicated to a single operating system, Linux, and have assembled smaller dedicated staffs of individuals who are committed to delivering a better solution. One such company is Penguin Computing.

Who is Penguin Computing?

Penguin Computing was founded in 1998 and is headquartered in San Francisco, with a staff of 65 employees. This will probably not come as a big surprise: **Penguin is dedicated to the Linux operating system. It is their 100% focus. They are a leading provider of scalable, robust, high-performance computer systems, cluster management software, and technical services.** There systems are turnkey – ready to use at installation, with exceptional support teams. Penguin is headed up by a management team led by Enrico Pesatori, formerly CEO of BlueArc and before that senior vice president and group general manager of Compaq's enterprise solutions and services group, and Don Becker, the original inventor of *Beowulf*, and founder of Scyld Software.

Penguin has a strong knowledge of Linux throughout their organization, especially in their pre-sales and post-sales support organizations. If you are going to name your company Penguin Computing, you had better have a total Linux solution, from manufacturing to support. Penguin does. As part of that solution, Penguin acquired the industry leader in Linux Clustering - Scyld Software - in August 2003. Scyld Beowulf provides complete clustering software, systems management framework, and highly scalable deployment platform as a robust solution for HPC clusters, based completely on standards, although currently no single standard for clustering has emerged. It is fully tested and packaged for ease of use operation in any data center – with or without a Linux guru on-hand. It is cross-platform; it will work on any commodity server with an Intel or AMD 32/64 bit CPU.

Their business is based upon building sustainable, repeatable business in the MSO and LSO Linux communities. Their emphasis is on building the confidence within their customer base to "convert" their audience from a **cult** to a "corporate **culture**". They are dedicated to making clustering an out of the box turnkey experience rather than an adventure. **Their goal is to make Linux the infrastructure for a mission-critical commercial application cluster for *Fortune 1000* companies,** not just another high-performance system for academia, government, and scientific applications such as biotech and geophysical applications**.**

They have a broad range of competitive, standards-based cluster products built around powerful and scalable commodity blade servers, rack-optimized for the enterprise. All of their systems are pre-configured at the factory with an emphasis on ease-of-use for HPC.

So far, over 2,000 organizations have recognized Penguin's edge in reducing the total cost of ownership of their information clusters. These customers range from academia to some of the larger names in the innovation of Internet Protocol-based networking technologies.

A Cluster in a Box

Penguin's legacy product set consists of a set of rack-optimized 32- and 64-bit server solutions, both AMD and Intel. Their AMD solutions are based upon the 64-bit *Opteron* processor, under the *Altus* label, while their Intel solutions are based upon 32- and 64-bit high performance *Xeon* servers, under the *Relion* label. Both sets of robust and reliable servers are available in 1U and 2U formats.

In addition to the rack-optimized servers, Penguin offers *BladeRunner*, a *Cluster-in-a Box* solution based upon dual 2.4 GHz 32-bit Intel *Xeon LV* Processors, with 64-bit *Xeon EM64T* and AMD *Opteron HE* processors on the roadmap for mid-2005 delivery. Scyld *BeoMaster* manages the clusters from the master node, with a single

point of management for all compute nodes.

BladeRunner Hardware Platform

A 4U BladeRunner chassis is an ideal server for data center consolidation as well as entry-level and departmental high performance applications. Capable of supporting up to 12 dual-processor blades (one master node, eleven compute nodes) for under $20K, BladeRunner can enable the IT staff to take advantage of up to 24 Intel Xeon LV processors, up to 48GB of ECC memory, with up to 4TB of storage. With reduced power and cooling requirements, reductions in cabling, integrated networking, and hardware component monitoring, the chassis provides an immediate return in reducing the TCO to the data center. BladeRunner is serviceable without the need for any tools and all active components have both redundancy and hot-swap capability, enabling the implementation of an HA environment to provide improved reliability. If HPC clustering is required, BladeRunner can be ordered with Ethernet switches, storage sub-systems, and cluster management software along with the Scyld Beowulf clustering software.

BladeRunner can also be pre-configured as an entry-level cluster with one master node and four compute nodes, along with a Scyld Beowulf license, priced under $15K. A full expansion cluster, with twelve-node cluster-in-a-box, is priced at under $25K.

Scyld Clustering Software

Release 29 of the Scyld Beowulf clustering software powers a Penguin cluster enabling a single system of management and unified process space for the entire cluster. Scyld Beowulf supports standard Linux tools and includes libraries and utilities specifically designed for clustering, resulting in a cluster that appears to be more like a traditional single multi-processor system. This reduces the cost of cluster application development, testing, training, and administration, further lowering the TCO for the data center. The major features of this release are:

- **Single Point Cluster Management** - for installation, configuration, administration,

and monitoring;

- **Unified Process Space** – Makes the cluster transparent to end users through a single process ID space with an SMP-like experience providing automatic process migration at job execution time;

- **Full Linux Distribution** – A fully standards based Red Hat implementation to maintain a familiar environment without the need to purchase additional Red Hat licenses; and

- **Integrated and Flexible HPC Tools** – Pre-configured and tested toolset for a complete HPC clustering solution.

Conclusion

Penguin Computing has emerged as a market leader in the supply of scalable Linux platforms for data center consolidation and the clustering of Linux applications. Penguin has taken ease of use for cluster processing to a new level with the acquisition and implementation of the Scyld software solution. **With a standards based hardware and software platform, Penguin has integrated a competitively priced, robust and scalable, high-performance computing system that can compete effectively with higher-priced (and proprietary) UNIX solutions while, at the same time, simplifying the data center infrastructure**.

With the new BladeRunner hardware and the Scyld clustering software, Penguin addresses the complexity of Linux clustering and provides a tested and documented platform for commercial applications, supported by a staff with a wealth of Linux expertise. With a proven solution and a broad customer base, Penguin has gained the attention of CIOs across the country. If your data center is concerned about the TCO of the enterprise-computing environment, perhaps you should take a look at Penguin Computing.

About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

➢ *The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.*

About the Author

David Reine is Director, Enterprise Systems for The Clipper Group. Mr. Reine specializes in enterprise servers, storage, and software, strategic business solutions, and trends in open systems architectures. He joined The Clipper Group after three decades in server and storage product marketing and program management for Groupe Bull, Zenith Data Systems, and Honeywell Information Systems. Mr. Reine earned a Bachelor of Arts degree from Tufts University, and an MBA from Northeastern University.

➢ *Reach David Reine via e-mail at dave.reine@clipper.com or at 781-235-0085 Ext. 123.*

Regarding Trademarks and Service Marks

The Clipper Group Navigator, **The Clipper Group Explorer**, **The Clipper Group Observer**, **The Clipper Group** *Captain's Log,* and *"clipper.com"* are trademarks of The Clipper Group, Inc., and the clipper ship drawings, *"Navigating Information Technology Horizons"*, and *"teraproductivity"* are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.